



EJERCICIO PRÁCTICO 8: ANOVA PARA MUESTRAS INDEPENDIENTES

CONTEXTO

Recordemos que la prueba T de Student permite comparar dos mediciones independientes de una variable aleatoria, que usualmente provienen diferentes grupos de individuos u objetos. El análisis de varianza (ANOVA), que nos permite comparar más de dos mediciones independientes, está compuesto de una etapa ómnibus más un análisis *post hoc* si corresponde, para comparar las medias de más de dos muestras.

OBJETIVOS DE APRENDIZAJE

1. Aplicar procedimientos de análisis de varianza para muestras independientes, usando el entorno R.
2. Aplicar herramientas disponibles en el entorno R para realizar un análisis *post-hoc* con muestras independientes.

ÉXITO DE LA ACTIVIDAD

1. El equipo obtiene y manipula correctamente los datos entregados y propone hipótesis pertinentes a contrastar para cada problema.
2. El equipo se asegura que cada caso cumple las condiciones para utilizar ANOVA con validez, usando gráficos o pruebas estadísticas auxiliares disponibles en el entorno R.
3. El equipo realiza de forma correcta y completa una prueba ANOVA ómnibus para cada problema.
4. El equipo determina correctamente si corresponde o no hacer un análisis *post-hoc* en cada caso, el que aplica de forma correcta y completa (cuando corresponda).
5. El equipo interpreta adecuadamente los resultados de las pruebas y responde las preguntas planteadas en cada caso, basándose en la prueba ómnibus y, si corresponde, considerando también el análisis *post-hoc*.

ACTIVIDADES

Un equipo de investigadores del área de interacción humano-información está estudiando si el área temática y el nivel de dificultad del problema de información influyen en el tiempo (en segundos) que toma un usuario en formular una consulta de búsqueda para resolver dicho problema. Para ello, han reclutado a un grupo de participantes voluntarios, asignados aleatoriamente a distintos grupos. Cada participante debe resolver tres problemas de información con diferentes niveles de dificultad: baja, media y alta. A su vez, cada grupo debe resolver problemas relacionados a una temática diferente. Los datos recolectados contemplan las siguientes variables:

- id: identificador único de cada participante.
 - area: área temática de los problemas que el participante debe responder. Variable categórica con los niveles Arquitectura, Biología, Computación, Economía, Física, Leyes, Literatura, Matemáticas, Música, Pedagogía, Psicología, Química.
 - dificultad: nivel de dificultad del problema resuelto. Variable categórica con los niveles Baja, Media y Alta.
 - tiempo: tiempo, en segundos, que toma al participante formular la consulta.
1. Copiar el enunciado del problema asignados como comentario de un script R.
 2. Descargar desde UVirtual el archivo EP08 Datos.csv con los datos a emplear.

3. Familiarizarse con los datos entregados, y enunciar las hipótesis nula y alternativa para el procedimiento ANOVA.
4. Analizar si se cumplen las condiciones para usar un procedimiento ANOVA y construir un script R para verificarlo.
5. Independiente del resultado anterior, aplicar una prueba ANOVA ómnibus a los datos y entregar una conclusión usando un nivel de significación adecuado.
6. Si corresponde, aplicar un análisis *post-hoc* e interpretar los resultados. En caso contrario, argumentar por qué no es necesario.
7. Redactar la respuesta a la pregunta planteada (comentario) en base a los resultados del análisis realizado.

Fuera del horario de clases, cada equipo debe subir el script realizado UVirtual con el nombre "EP08-respuesta-grupo-i", donde i es el número de grupo asignado. Las respuestas deben subirse antes de las 23:30 del sábado 23 de abril.

PREGUNTAS

Grupo 1:

En este momento, los investigadores buscan determinar si existen diferencias en el tiempo que tardan los usuarios en formular una consulta para un problema de dificultad fácil en las áreas de computación, literatura y química.

Grupo 2:

En este momento, los investigadores buscan determinar si existen diferencias en el tiempo que tardan los usuarios en formular una consulta para un problema de dificultad difícil en las áreas de psicología, música y matemáticas.

Grupo 3:

En este momento, los investigadores buscan determinar si existen diferencias en el tiempo que tardan los usuarios en formular una consulta para un problema de dificultad media en las áreas de economía, literatura y arquitectura.

Grupo 4:

En este momento, los investigadores buscan determinar si existen diferencias en el tiempo que tardan los usuarios en formular una consulta para un problema de dificultad media en las áreas de leyes, música y matemáticas.

Grupo 5:

En este momento, los investigadores buscan determinar si existen diferencias en el tiempo que tardan los usuarios en formular una consulta para un problema de dificultad fácil en las áreas de biología, economía y computación.

Grupo 6:

En este momento, los investigadores buscan determinar si existen diferencias en el tiempo que tardan los usuarios en formular una consulta para un problema de dificultad difícil en las áreas de economía, psicología y arquitectura.

Grupo 7:

En este momento, los investigadores buscan determinar si existen diferencias en el tiempo que tardan los usuarios en formular una consulta para un problema de dificultad fácil en las áreas de biología, leyes y psicología.

Grupo 8:

En este momento, los investigadores buscan determinar si existen diferencias en el tiempo que tardan los usuarios en formular una consulta para un problema de dificultad difícil en las áreas de física, arquitectura y música.

Grupo 9:

En este momento, los investigadores buscan determinar si existen diferencias en el tiempo que tardan los usuarios en formular una consulta para un problema de dificultad media en las áreas de economía, biología y literatura.

Grupo 10:

En este momento, los investigadores buscan determinar si existen diferencias en el tiempo que tardan los usuarios en formular una consulta para un problema de dificultad difícil en las áreas de leyes, química y economía.

CRITERIOS DE EVALUACIÓN

Categoría	Nivel de logro	Puntaje
Datos	Obtienen los datos necesarios para realizar la prueba solicitada correctamente, tanto en formato ancho como en formato largo	2
Formulación de hipótesis	Formula con claridad y explícitamente hipótesis nula y alternativa que mencionan las medias de las diferencias de las mediciones independientes.	3
Condiciones	Verifica el cumplimiento de todas las condiciones requeridas por ANOVA para muestras independientes, usando para ello gráficos o pruebas estadísticas adecuadas.	3
Prueba ómnibus	Realiza una prueba ANOVA omnibus para muestras independientes, usando para ello un formato pertinente para los datos y los parámetros correspondientes.	3
Prueba post hoc	Realiza una prueba post-hoc para muestras independientes, usando para ello un formato pertinente para los datos y los parámetros correspondientes; o bien señala que no corresponde efectuar un procedimiento post-hoc, argumentando de manera correcta y completa.	3
Conclusión	Entrega una conclusión correcta y completa a la pregunta planteada, basándose en el resultado de la prueba realizada y el contexto del problema.	3
Código fuente	Escribe código R -ordenado, bien indentado, sin sentencias espurias y bien comentado- que realiza de forma completa y correcta la prueba seleccionada con los datos adecuados en cada caso	3
Ortografía y redacción	Escribe con buena ortografía y redacción, usando vocabulario propio de la disciplina y el contexto del problema.	3
Puntaje total		23